AD-A049 510 MASSACHUSETTS INST OF TECH CAMBRIDGE DEPT OF MATHEMATICS F/G 22/3

A SATELLITE CONTROL PROBLEM.(U)

DEC 77 H CHERNOFF, A J PETKAU N00014-75-C-0555

UNCLASSIFIED TR-9 NL

1 of 1

ADA049510

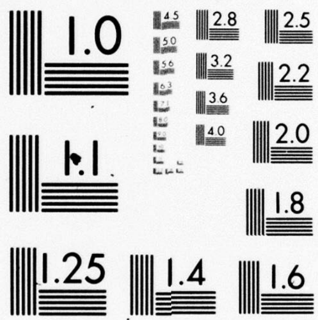
END

DATE

FILMED

3-78

DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3

A SATELLITE CONTROL PROBLEM

BY

HERMAN CHERNOFF

A. JOHN PETKAU (University of British Columbia, CANADA)

TECHNICAL REPORT NO. 9

DECEMBER 22, 1977

PREPARED UNDER CONTRACT
N00014-75-C-0555 (NR-042-331)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

6

A SATELLITE CONTROL PROBLEM

by

10 Herman Chernoff

A. John Petkau (University of British Columbia, CANADA)

9 Technical Report No. 9

11 22 Dec [redacted] 77

12 56p.

14 TR-9

Prepared Under Contract

15 N00014-75-C-0555 (NR-042-331)

For the Office of Naval Research

DEPARTMENT OF MATHEMATICS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

CAMBRIDGE, MASSACHUSETTS

220 021

1. Introduction

It is desired to keep a satellite close to a fixed point in space when it is subject to random forces. Fuel may be used to accelerate it at appropriate times. The permitted acceleration is bounded. While it may be desirable to maximize the performance for a given amount of fuel, we shall consider the fuel as available in unlimited amounts at a fixed cost per unit. Our object will be to minimize the long run expected average cost per unit time assuming some cost for being away from target. The random forces are modeled by Brownian Motion. In this discussion we shall treat the position in space as one-dimensional and given by x_1 and the cost of being at x_1 is x_1^2 per unit time. Let x_2 = velocity and $\underline{x} = \underline{x}(t) = (x_1(t), x_2(t))$ describes the current state of the satellite which is subject to the laws

$$(1.1) \qquad dx_1 = x_2 dt$$

$$(1.2) \qquad dx_2 = u dt + \sigma dw(t) \qquad |u| \le u_0$$

where $w(t)$ represents standard Brownian Motion with $E[dw(t)]=0$ and $E[dw^2(t)] = dt$ and where u represents the acceleration which is subject to control depending on t and past history.

The running cost per unit time is given by $r(\underline{x},u)$ where

$$(1.3) \qquad dR(t) = (c|u| + x_1^{\,2})dt = r(\underline{x},u)dt$$

We shall seek a policy $\mathcal{P}$ which minimizes

$$(1.4) \qquad \gamma = \lim_{T\to\infty} \sup E\{C(\underline{x},0,T)/T\} \quad .$$

where $C(\underline{x},t,T)$ is the total cumulated cost over the time period (t,T) when the state process $\underline{X}(t'): t \leq t' \leq T$ originates at $\underline{X}(t) = \underline{x}$ at time t. The random process $\underline{X}(t')$ is dependent on the policy and hence the expectation involves $\mathcal{P}$ implicitly. When there is a possibility of confusion, the policy $\mathcal{P}$ will be indicated as a superscript on the expectation, e.g. $E^{\mathcal{P}}C(\underline{x},t,T)$.

Intuitive considerations show that using an acceleration of u for a time interval dt and 0 for dt is equivalent to using $u/2$ for $2dt$. Thus the search for an optimal policy may reasonably be confined to policies which use only the values $u = \pm u_0$ and 0. For a stationary policy where $u = u(x_1,x_2)$, independent of t and past history (with some minor abuse of notation), the policy can be described by dividing up the (x_1,x_2) state space into three regions A_+, A_- and A_0 where $u = +u_0$, $-u_0$ and 0 respectively.

The main object of this paper is to describe a simple numerical approach for deriving and evaluating an optimal policy. The basic method is to apply backward induction to the Markov Decision problem that is formed from approximating the continuous space time problem described above by a bounded discrete space, discrete time version of the problem. A similar approach was applied by Kushner and Kleinman [6, 7, 8]. The main difference in this approach is in the method of handling the edge effects at the boundary of the bounded space.

Since the method is iterative and uses an initial approximation, this was obtained from an approximation to the solution of the deterministic version of the original problem where there are no random forces. In Section 2 the cost associated with a special suboptimal policy in the deterministic problem is evaluated and optimality conditions are introduced to explain how this candidate was selected and why it is suboptimal. In Section 3 the optimal policy for the deterministic problem is described.

For both the deterministic and stochastic versions of the problem, the homogeneity of the cost functions permit one to standardize the problem and effectively eliminate two of the parameters c, u_0, and σ by applying linear transformations to the variables x_1, x_2, t, u and w. These transformations are described in Section 4.

After discussion of the relationship between the solution and bounds on the solution and a free boundary problem in Section 5 the discrete approximation to the problem is described in Sections 6, 7 and 8. Sections 9 and 10 are devoted to several alternative treatments of edge effects. Some miscellaneous remarks appear in Section 11 and finally Section 12 presents some results of preliminary computations.

2. The Deterministic Version - A Suboptimal Policy

If there is no random noise, i.e., $\sigma = 0$, it is easy to find control policies which bring the satellite to $x_1 = 0$ with zero velocity in a finite time interval. Thus, for the deterministic version of the problem it is reasonable to consider simply the total cumulated cost $V(x_1, x_2)$ associated with a given policy and to minimize that.

It is clear that an optimal policy for the deterministic problem will correspond to decomposing the state space into three sets A_+, A_0, A_- on which $u = u_0$, 0 and $-u_0$ respectively.

For one of our numerical procedures for the stochastic problem we shall make use of an approximation to the optimal solution of the deterministic problem. The precision of our approximation is not crucial and so we shall use a moderately convenient approximation. In this section we describe that suboptimal procedure, and compute its cumulated cost $V(x_1, x_2)$. The source of this policy as well as an explanation of why it

is suboptimal in terms of conditions of optimality will conclude this section. We shall follow this in Section 3 by a brief description of the optimal policy for the special values of the parameters $u_0 = c = 1$.

The suboptimal policy is described in terms of three sets A_+, A_0, and A_- where one applies $u = +u_0$, 0, and $-u_0$ respectively (See Figure 1). These in turn may be expressed in terms of two curves C_0 and C^* and their reflections about the origin

(2.1) $$C_0 = \{(x_1,x_2) : x_1 = x_2^2/(2u_0), x_2 \leq 0\}$$

(2.2) $$C^* = \{(x_1,x_2) : x_1^3 = 6cx_2^2 + x_2^6/(8u_0^3), \quad x_2 \leq 0\}$$

The set A_0 consists of the region between C_0 and C^* and its reflection. More precisely $(x_1,x_2) \in A_0$ if $x_2 \leq 0$ and $x_2^2/(2u_0) \leq x_1 \leq [6cx_2^2 + x_2^6/(8u_0^3)]^{1/3}$ or if $(-x_1,-x_2)$ satisfies these two inequalities. The set A_- consists of the region between C^* and the reflection of C_0 and A_+ is the remaining part of the plane.

Under this policy and the laws of motion

(2.3) $$dx_1 = x_2 dt$$

(2.4) $$dx_2 = u\,dt$$

a satellite whose state (x_1,x_2) is in A_- will move to C^* under $u = -u_0$ following a parabolic path where $x_1 + x_2^2/(2u_0)$ is constant. From C^* it will move to C_0 keeping x_2 fixed. Once it hits the parabolic path C_0, $u = u_0$ and it follows C_0 to the origin. We compute V for this policy by retracing the path of the satellite from the origin to (x_1,x_2).

If $(x_1,x_2) \varepsilon C_0$,

$$V(x_1,x_2) = V^{(0)}(x_1,x_2) := \int_{\underline{x}}^{0} (cu_0 + x_1'^2)dt' = \int_{x_2}^{0} (cu_0 + \frac{x_2'^4}{4u_0^2})\frac{dx_2'}{u_0}$$

(The first integral is a line integral along C_0 from $\underline{x} = (x_1,x_2)$ to $(0,0)$.)

$$\text{(2.5)} \qquad V^{(0)}(x_1,x_2) = -cx_2 - x_2^5/(20u_0^3)$$

If $(x_1,x_2) \varepsilon A_0$ with $x_2 < 0$,

$$V(x_1,x_2) = V^{(1)}(x_1,x_2) := V^{(0)}(x_{10},x_{20}) + \int_{\underline{x}}^{\underline{x}_0} x_1'^2 \, dt'.$$

where $\underline{x}_0 = (x_{10},x_{20})$ is the state at which the satellite originating at $(x_1,x_2) \varepsilon A_0$ first intersects C_0. Then $x_{10} = x_{20}^2/(2u_0)$, $x_{20} = x_2$, and

$$\int_{\underline{x}}^{\underline{x}_0} x_1'^2 dt = \int_{x_1}^{x_{10}} (x_1')^2 \frac{dx_1'}{x_2} = \frac{x_{10}^3 - x_1^3}{3x_2} = \frac{x_2^5}{24u_0^3} - \frac{x_1^3}{3x_2}$$

$$(2.6) \qquad V^{(1)}(x_1, x_2) = -cx_2 - \frac{x_2^5}{120u_0^3} - \frac{x_1^3}{3x_2}$$

If $(x_1, x_2) \varepsilon A_-$,

$$V(x_1, x_2) = V^{(2)}(x_1, x_2) := V^{(1)}(x_1^*, x_2^*) + \int_{\underline{x}}^{\underline{x}^*} (cu_0 + x_1'^2) dt'$$

where $\underline{x}^*$ is the point on C^* which the satellite first reaches from $\underline{x} \varepsilon A_-$. Then $\underline{x}^*$ is determined by $x_1^{*3} = 6cx_2^2 + x_2^{*6}/(8u_0^3)$ and $x_1^* + x_2^{*2}/(2u_0) = a := x_1 + x_2^2/(2u_0)$. Then

$$\int_{\underline{x}}^{\underline{x}^*} (cu_0 + x_1'^2) dt' = \int_{x_2}^{x_2^*} [cu_0 + (a - \frac{x_2'^2}{2u_0})^2] \frac{dx_2'}{-u_0}$$

$$= (c + \frac{a^2}{u_0})(x_2 - x_2^*) - \frac{a}{3u_0^2}(x_2^3 - x_2^{*3}) + \frac{(x_2^5 - x_2^{*5})}{20u_0^3}$$

$$(2.7) \quad V^{(2)}(x_1, x_2) = -\frac{x_1^{*3}}{3x_2^*} - cx_2^* - \frac{x_2^{*5}}{120u_0^3} + (c + \frac{a_2}{u_0})(x_2 - x_2^*)$$

$$- \frac{a}{3u_0^2}(x_2^3 - x_2^{*3}) + \frac{(x_2^5 - x_2^{*5})}{20u_0^3}$$

For $(x_1,x_2) \varepsilon A_+$ and that part of A_0 where $x_2 > 0$, we may apply symmetry to obtain $V(x_1,x_2) = V(-x_1,-x_2)$.

Having computed V, we may ask how C^* was selected and why $\dot{V}$ is not optimal. First we deal with conditions for optimality.

Let $\mathcal{P}$ be a policy which assigns control value $u = u(x_1,x_2,t)$. Let $V(x_1,x_2,t)$ be the cumulated cost associated with $\mathcal{P}$ from time t on if $\underline{x}(t) = (x_1,x_2)$. Let

$$(2.8) \qquad \mathcal{L}_0 V = x_2 V_{x_1} + u V_{x_2}$$

Note that $\mathcal{L}_0$ depends on $\mathcal{P}$ since it involves the control u. If $V < \infty$, the laws of motion (2.3), (2.4) imply that along the path prescribed by $\mathcal{P}$, V changes by $dV = \mathcal{L}_0 V dt = (x_2 V_{x_1} + uV_{x_2})dt$ and hence

$$(2.9) \qquad x_2 V_{x_1} + uV_{x_2} + c|u| + x_1^2 = \mathcal{L}_0 V + r(\underline{x},u) = 0$$

which may be interpreted as

$$(2.9a) \qquad dV + dR = 0,$$

where $dR = (c|u| + x_1^2)dt = r(\underline{x},u)dt$ and r is the running cost. For a policy defined by A_+, A_0, and A_- as the one considered above, u assumes the values u_0, 0 and $-u_0$

respectively on these sets. For a stationary or time independent policy where $u=u(\underline{x},t)$ is independent of t, V is independent of t.

Suppose now that $V=V(x_1,x_2)$ is a function vanishing at the origin and with the property that no matter what policy is followed,

$$dV + dR \geq 0. \tag{2.10a}$$

Integrating along the path followed by a policy $\mathcal{P}^*$ which drives the satellite from (x_1,x_2) to $(0,0)$ we have

$$V(0,0) - V(x_1,x_2) + \int dR = V(0,0) - V(x_1,x_2) + V^* \geq 0$$

where V^* is the cumulated cost for $\mathcal{P}^*$. Since $V(0,0)=0$,

$$V^* \geq V(x_1,x_2) \tag{2.11}$$

If there is also a policy $\mathcal{P}$ for which $dV + dR = 0$ it would follow that V is the cost for $\mathcal{P}$ and $\mathcal{P}$ is optimal. Thus the optimality of $\mathcal{P}$ would follow <u>if we could show in addition to</u> (2.9)

$$x_2V_{x_1} + uV_{x_2} + c|u| + x_1^{\,2} \geq 0 \tag{2.10}$$

for all $|u| \leq u_0$ and all x.

In effect we have proved

Theorem 2.1 For any policy $\mathcal{P}$ with finite cummulated cost, $dV + dR = 0$. If V is the cumulated cost associated with a stationary policy $\mathcal{P}$ and $dV + dR \geq 0$ for each policy $\mathcal{P}^*$ then $\mathcal{P}$ is optimal.

In our application we can combine equations (2.9) and (2.10) and we see that optimality of a policy defined by A_+, A_0 and A_- requires

$$V_{x_2} \leq -c \quad \text{on } A_+$$

$$|V_{x_2}| \leq c \quad \text{on } A_0 \tag{2.12}$$

$$V_{x_2} \geq c \quad \text{on } A_-$$

We now return to the policy proposed at the beginning of this section. Given our decision to apply $u=u_0$ on C_0 and to include in A_0 points in the fourth quadrant immediately to the right of C_0, we have $V=V^{(1)}$ for those points. But for such points $|V^{(1)}_{x_2}| = |-c-x_2{}^4/24u_0{}^3 + x_1{}^3/3x_2| \leq c$ as long as $x_1{}^3 \geq x_2{}^6/(8u_0{}^3)$ and $x_1{}^3 \leq 6cx_2{}^2 + x_2{}^6/(8u_0{}^3)$. Thus, given our decision to apply $u = u_0$ on C_0, the choice of C^* for the boundary of A_- is optimal.

However, if we study V_{x_2} on the A_-, as we retrace the path of a satellite from C^* along the parabola

$$x_1 + x_2^2/(2u_0) = a = x_1^* + x_2^{*2}/(2u_0)$$

it is possible to see, by calculations similar to that to be given in Section 3, that as x_2 increases from x_2^*, V_2 increases from c at first <u>but</u> eventually decreases below c in part of A_- in the second quadrant. But optimality would demand that that part of A_- should be in A_0 and our policy fails to satisfy the optimality conditions (2.12).

An intuitive explanation is that by using $u = u_0$ on C_0, we slow down the satellite while x_1 is large and accumulate a large cost by staying in a region of large x_1 too long. Apparently it is preferable to pass through C_0 before slowing down. While this means we overshoot the $x_1 = 0$ target, we do so where x_1 is relatively small and the additional cost incurred from having to retrace our path is less than that of tarrying too long in a region of large x_1.

Although the policy of this section is suboptimal, it resembles the optimal policy sufficiently to serve as a useful device for the numerical analysis of the stochastic problem.

It is of some interest to repeat that the optimality condition is violated when $x_2 > 0$ only if $x_1 < 0$. Hence

this policy is optimal if we add the restriction that $x_1 > 0$. Thus it represents a deterministic solution to a problem of a soft landing on a planet from a rocket stationed vertically over the point of impact. Of course the force of gravity must be assumed to be constant and this particular solution is meaningful only for a cost function which is rather peculiar in the soft landing application. Shepp studied a similar deterministic problem where the cost was that of fuel and the time to reach the target. In that problem, the optimal policy does use C_0 but C^* is replaced by $\{\underline{x}^* : x_1^* = (1+4cu_0)x_2^{*2}/(2u_0)\}$

3. The Deterministic Version - Optimal Policy

In the interest of simplicity let us consider the deterministic problem for the case $c = u_0 = 1$. There is no real loss of generality since a linear transformation of the parameters and state variables to be described in Section 4, permits us to normalize our problem to this standard case.

Heuristic considerations suggest that the optimal policy may be described by A_0, A_- and A_+ bounded by new curves C_0, C^* in the fourth quadrant and their reflections about the origin. The cost associated with the optimal policy $V(x_1,x_2)$ will satisfy

$$x_2V_{x_1} + uV_{x_2} + |u| + x_1^2 = 0$$

with $u = 0, -1$, and $+1$ on A_0, A_-, and A_+ respectively. Moreover, the optimality conditions (2.12) require $|V_{x_2}| \leq 1$ on A_0, $V_{x_2} \geq 1$ on A_- and $V_{x_2} \leq -1$ on A_+. Given a point (x_{10}, x_{20}) on C_0 where $V_{x_2} = -1$ and $x_2 V_{x_1} = -x_1^2$, we compute V backwards from this point along the path of points going to C_0

$$V(x_1,x_2) = V(x_{10},x_{20}) + \int_{\underline{x}}^{\underline{x}_0} x_1'^2 \, dt'$$

$$= V(x_{10},x_{20}) + (x_{10}^3 - x_1^3)/3x_2$$

where $x_{20} = x_2$.

$$V_{x_2}(x_1,x_2) = \frac{\partial V(x_{10},x_{20})}{\partial x_{10}} \frac{\partial x_{10}}{\partial x_2} + \frac{\partial V(x_{10},x_{20})}{\partial x_{20}} + \frac{x_{10}^2}{x_2} \frac{\partial x_{10}}{\partial x_2} - \frac{(x_{10}^3 - x_1^3)}{3x_2^2}$$

But since $x_2 V_{x_1} + x_1^2 = 0$ on A_0 and $V_{x_2} = -1$ on C_0

$$V_{x_2}(x_1,x_2) = -1 - (x_{10}^3 - x_1^3)/3x_2^2.$$

As we retrace the paths from (x_{10}, x_{20}), x_2 remains fixed at x_{20} and x_1 increases. Thus V_{x_2} increases from -1 to $+1$ when $x_1^3 = x_{10}^3 + 6x_{20}^2$. Thus the boundary C^*, determined by the optimality conditions and C_0, is

(3.1) $$C^* = \{\underline{x}^* : x_1^{*3} = x_{10}^{\ 3} + 6x_{20}^{\ 2},\ x_2^* = x_{20}\}$$

Given a point (x_1^*, x_2^*) on C^* we compute V for points on the path $x_1 + x_2^{\ 2}/2 = a = x_1^* + x_2^{*2}/2$ which leads to (x_1^*, x_2^*)

$$V(x_1,x_2) = V(x_1^*,x_2^*) + \int_{\underline{x}}^{\underline{x}^*} (1+x_1'^2)dt'$$

$$= V(x_1^*,x_2^*) + \int_{x_2}^{x_2^*} [1+(a-\frac{x_2'^2}{2})^2]\ \frac{dx_2'}{(-1)}$$

$$= V(x_1^*,x_2^*) + (1+a^2)(x_2-x_2^*) - \frac{a}{3}(x_2^{\ 3}-x_2^{*3}) + \frac{x_2^{\ 5}-x_2^{*5}}{20}$$

$$V_{x_2} = \frac{\partial V(x_1^*,x_2^*)}{\partial x_1^*}\ \frac{\partial x_1^*}{\partial x_2} + \frac{\partial V(x_1^*,x_2^*)}{\partial x_2^*}\ \frac{\partial x_2^*}{\partial x_2} + [1+(a-\frac{x_2^{\ 2}}{2})]$$

$$- [1+(a-\frac{x_2^{*2}}{2})^2]\ \frac{\partial x_2^*}{\partial x_2} + [2a(x_2-x_2^*) - \frac{1}{3}(x_2^{\ 3} - x_2^{*3})]\ \frac{\partial a}{\partial x_2}$$

Substituting $\partial V(x_1^*,x_2^*)/\partial x_1^* = -x_1^{*2}/x_2^*$, $\partial V(x_1^*,x_2^*)/\partial x_2^* = 1$, $a = x_1^* + x_2^{*2}/2 = x_1 + x_2^{\ 2}/2$, $\partial a/\partial x_2 = \partial x_1^*/\partial x_2 + x_2^*\partial x_2^*/\partial x_2 = x_2$, we have

$$V_{x_2} = \frac{-x_1^{*2}}{x_2^{*}} \left[\frac{\partial x_1^{*}}{\partial x_2} + x_2^{*} \frac{\partial x_2^{*}}{\partial x_2} \right] + 1 + \left(a - \frac{x_2^{2}}{2}\right)^2$$

$$+ \left[2a(x_2 - x_2^{*}) - \frac{1}{3}(x_2^{3} - x_2^{*3})\right]x_2$$

$$(3.2) \quad V_{x_2} = H(x_2) := 1 - \frac{x_1^{*2}}{x_2^{*}} x_2 + \left(a - \frac{x_2^{2}}{2}\right)^2 + x_2\left[2a(x_2 - x_2^{*})\right.$$

$$\left. - \frac{1}{3}(x_2^{3} - x_2^{*3})\right]$$

As we retrace the path from x_1^*, x_2^*, a remains fixed but x_2 increases. As a polynomial in x_2, H increases from 1 at $x_2 = x_2^*$ but eventually decreases again since the coefficient of x_2^4 is $(1/4) - (1/3) = -1/12$. Let $\tilde{C}$ be the curve of $(\tilde{x}_1, \tilde{x}_2)$ for which $\tilde{x}_2$ is the first $x_2 > x_2^*$ where $H(x_2) = 1$ and $\tilde{x}_1 + \tilde{x}_2^2/2 = a$. It is easy to see that $\tilde{C}$ is in the second quadrant since

$$H'(x_2) = -(x_1^{*2}/x_2^{*}) + 2a(x_2 - x_2^{*}) - (x_2^{3} - x_2^{*3})/3$$

is positive at $x_2 = x_2^* < 0$ and

$$H''(x_2) = 2a - x_2^{2} = 2x_1$$

is positive as long as x_1 is positive. Thus $H(x_2) > 1$ for

(x_1, x_2) in the fourth quadrant above C^* and in the first quadrant.

Clearly $\tilde{C}$ serves as a reflection of C_0 to yield a new boundary point of A_0. Thus our regions and boundaries are determined by the procedure of going from (x_{10}, x_{20}) to (x_1^*, x_2^*) and then to $(\tilde{x}_1, \tilde{x}_2)$ described above.

A relatively simple asymptotic analysis using (3.1) and $H(\tilde{x}_2) = 1$ and the fact that $0 < x_{10} < x_{20}^2/2$, shows that for very small negative x_2, the corresponding values of x_{10} on C_0 and x_1^* on C^* are given by

$$(3.3)\qquad \begin{aligned} x_{10} &\approx x_2^2/2 \\ x_1^* &\approx 6^{1/3} x_2^{2/3} \end{aligned}$$

Also the corresponding point $(\tilde{x}_1, \tilde{x}_2)$ satisfies

$$(3.4)\qquad \begin{aligned} \tilde{x}_2 &\approx -3^{5/9} 2^{8/9} x_2^{1/9} \approx 3.409\, x_2^{1/9} \\ \tilde{x}_1 &\approx -\tilde{x}_2^2/2 \approx -5.811\, x_2^{2/9} \end{aligned}$$

For large x_2, the corresponding values of x_{10} and x_1^* are given by

$$x_{10} \approx 0.44462\, x_2^{\,2}, \quad \sqrt{2}x_{10} \approx 0.94300 x_2$$

(3.5)

$$x_1^* \simeq 0.44462\, x_2^{\,2}, \quad \sqrt{2}x_1^* \approx 0.94300\, x_2.$$

Moreover

$$\tilde{x}_2 \simeq -4.13016\, x_2$$

(3.6)

$$\tilde{x}_1 \approx -7.58449\, x_2^{\,2}.$$

Applying (3.1), $x_1^* = (x_{10}^3 + 6x_2^{\,2})^{1/3} \approx x_{10}(1+2x_2^{\,2}x_{10}^{-3})$ and we have

$$\text{(3.7)} \quad x_1^* - x_{10} \simeq 10.11685\, x_2^{\,2}.$$

Figure 1 shows the optimal region A_0 obtained by starting from $\underline{x}_0$ with $x_{10} = x_{20}^{\,2}/2$ for small x_{20} and computing a sequence of successive values of $\underline{x}^*$ and $\underline{x}_0' = -\underline{\tilde{x}}$. The same calculation with initial point $x_{10} = 0$ leads to almost identical points when the initial x_{20} is small.

4. Transformations.

The homogeneous nature of the cost $x_1^{\,2}$ permits one to normalize both the stochastic and deterministic versions of the problem by means of simple linear transformations. This normalization effectively reduces the number of parameters that need to be

considered by two and is of considerable convenience although not of fundamental importance.

We start with the deterministic problem. Let

(4.1) $x_1^* = a_1 x_1, x_2^* = a_2 x_2,\ t^* = a_3 t,\ u^* = a_4 u.$

Then applying (1.1)-(1.3) we have,

$$dx_1^* = a_1 dx_1 = a_1 a_2^{-1} a_3^{-1} x_2^* dt^*$$

$$dx_2^* = a_2 u dt = a_2 a_4^{-1} a_3^{-1} u^* dt^*, \quad |u^*| \leq a_4 u_0$$

$$dR = (c a_4^{-1} a_3^{-1} u^* + a_3^{-1} a_1^{-2} x_1^{*2}) dt^*$$

$$= a_1^{-2} a_3^{-1} [c a_4^{-1} a_1^{2} u^* + x_1^{*2}]\ dt = a_1^{-2} a_3^{-1} dR^*.$$

If we set $a_1 a_2^{-1} a_3^{-1} = 1$, $a_2 a_4^{-1} a_3^{-1} = 1$, $u_0^* = a_4 u_0$, and $c^* = c a_4^{-1} a_1^{2}$ we have

(4.2)
$$a_1 = (c^* u_0^* / c u_0)^{1/2}, \quad a_2 = (c^* u_0^{*3} / c u_0^{3})^{1/4}$$
$$a_3 = (c^* u_0 / c u_0^*)^{1/4}, \quad a_4 = u_0^* / u_0$$

and our problem is now in the original form except that c and u_0 have been replaced by c^* and u_0^* and the cost

(4.3) $R(x_1,x_2;c,u_0) = (c^5 u_0{}^3/c^{*5} u_0{}^{*3})^{1/4} R^*(a_1x_1,a_2x_2;c^*,u_0{}^*)$

If we wish we can normalize the starred version by setting $c^* = u_0{}^* = 1$ in which case the solution of the original problem can be expressed in terms of that of the normalized one.

The stochastic version of the problem is a little more complicated. Here we apply

(4.4) $x_1{}^* = a_1x_1,\ x_2{}^* = a_2x_2,\ t^* = a_3t_0,\ u^* = a_4u,\ w^* = a_5w$

to (1.1)-(1.3). Proceeding as before we have

$$dx_1{}^* = a_1a_2{}^{-1}a_3{}^{-1}x_2{}^*dt^*$$

$$dx_2{}^* = a_2a_3{}^{-1}a_4{}^{-1}u^*dt^* + a_2\sigma a_5{}^{-1}\,dw^*\ ,\ |u^*|\le a_4u_0$$

$$dR = ca_3{}^{-1}a_4{}^{-1}|u^*|dt^* + a_1{}^{-2}a_3{}^{-1}x_1{}^{*2}dt^*$$

$$= a_1{}^{-2}a_3{}^{-1}[ca_1{}^2a_4{}^{-1}|u^*|dt^* + x_1{}^{*2}dt^*]$$

and

$$E(dw^*)^2 = a_5{}^2dt = a_5{}^2a_3{}^{-1}dt^*$$

Our problem is left invariant except for the transformation of u_0, σ, and c to u_0^*, σ_0^*, and c^* if $a_1 a_2^{-1} a_3^{-1} = 1$, $a_2 a_3^{-1} a_4^{-1} = 1$, $a_2 a_5^{-1} \sigma = \sigma^*$, $a_4 u_0 = u_0^*$, $a_5^2 a_3^{-1} = 1$, and $c^* = c a_1^2 a_4^{-1}$. Thus, for given u_0^*, σ^* select

$$a_1 = \sigma^{*4} u_0^3 / \sigma^4 u_0^{*3}, \qquad a_2 = \sigma^{*2} u_0 / \sigma^2 u_0^*$$

$$a_3 = \sigma^{*2} u_0^2 / \sigma^2 u_0^{*2}, \qquad a_4 = u_0^* / u_0 \tag{4.5}$$

$$a_5 = \sigma^* u_0 / \sigma u_0^*, \qquad c^* = c \sigma^{*8} u_0^7 / \sigma^8 u_0^{*7}.$$

Finally the cumulated cost $R(u_0, \sigma, c, t)$ over the time period $(0,t)$ is transformed by

$$R(u_0, \sigma, c, t) = (\sigma^{10} u_0^{*8} / \sigma^{*10} u_0^8) R(u_0^*, \sigma^*, c^*, t^*) \tag{4.6}$$

As an illustration to which we shall refer later, if $u_0 = \sigma = c = 1$ and we set $u_0^* = 1$ and $\sigma^* = 2$, then $a_1 = 16$, $a_2 = 4$, $a_3 = 4$, $a_4 = 1$, $a_5 = 2$, and $c^* = 256$.

5. <u>Stochastic Control Problem and Free Boundary Problem.</u>

Given a policy $\mathcal{P}$ and an initial value $x = \underline{X}(t_0)$ of the state at time t_0, the state $\underline{X}(t)$ at time t has a corresponding probability distribution. The cumulated cost over

$(t_0, t_1]$ is given by

$$(5.1) \quad C(\underline{x}, t_0, t_1) = \int_{t_0}^{t_1} dR(t) = \int_{t_0}^{t_1} r[\underline{X}(t), u(t)]dt$$

The heuristic assumption that for a stationary policy

$$(5.2) \quad E^{\mathcal{P}}\{C(\underline{x}, t_0, t_1)\} = \gamma(t_1 - t_0) + v(\underline{x}) + o(1)$$

as $t_1 - t_0 \to \infty$ together with $C(\underline{x}, t, t_1) = dR + C(\underline{X}(t+dt), t+dt, t_1)$ suggests

$$(5.3) \quad E^{\mathcal{P}}\{dv + dR\} = \gamma\, dt$$

where

$$E^{\mathcal{P}}\{dv\} = E^{\mathcal{P}}\{v[\underline{X}(t+dt)] \mid \underline{X}(t) = \underline{x}\} - v(\underline{x})$$

$$= [x_2 v_{x_1} + u v_{x_2} + \frac{\sigma^2}{2} v_{x_2 x_2}]\, dt$$

or

$$(5.3') \quad x_2 v_{x_1} + u v_{x_2} + \frac{\sigma^2}{2} v_{x_2 x_2} + c|u| + x_1^2 = \gamma .$$

If we define

$$(5.4)\quad \mathcal{L}v = x_2 v_{x_1} + u v_{x_2} + \frac{\sigma^2}{2} v_{x_2 x_2}$$

then (5.3') may be written as

$$(5.3")\quad \mathcal{L}v + r(\underline{x},u) = \gamma$$

Bather [1] called the function v, first introduced by Howard [5], the potential function. It has also been called the value difference function. Thus if $\mathcal{P}$ is a stationary policy which imposes $u = 0$, $\pm u_0$ in A_0, A_+, A_-, Equation (5.3') converts into separate equations in each region. The heuristic reasoning leads to a more solid interpretation if we introduce a truncated version of our problem which terminates at time t_1 with terminal cost $v[\underline{X}(t_1)]$. The expected cost of $\mathcal{P}$ for this problem is

$$(5.5)\quad D_v^{\mathcal{P}}(\underline{x},t_0,t_1) = E^{\mathcal{P}} C(\underline{x},t_0,t_1) + E^{\mathcal{P}}\{v[\underline{X}(t_1)] \mid \underline{X}(t_0) = \underline{x}\}$$

If v is a function which satisfies (5.3) then integrating (5.3) (formally, this is an application of Dynkin's formula [3, p.133]) gives

(5.6) $D^{\mathcal{P}}_{v}(\underline{x},t_0,t_1) = \gamma(t_1-t_0) + v(\underline{x})$

and γ is the expected long run average cost of $\mathcal{P}$.

Furthermore if $\mathcal{P}$ is a stationary policy and v is a function such that

(5.7) $E^{\mathcal{P}^*}\{dv + dR\} \geq \gamma\, dt = E^{\mathcal{P}}\{dv + dR\}$

for all x and all policies $\mathcal{P}^*$, then

$D_v^{\mathcal{P}^*}(\underline{x},t_0,t_1) \geq \gamma(t_1-t_0) + v(\underline{x}) = D_v(\underline{x},t_0,t_1)$ and $\mathcal{P}$ is optimal for the truncated problem with terminal cost v. If $\mathcal{P}^*$ is a stationary policy for which $E^{\mathcal{P}^*}\{v[\underline{X}(t_1)] \mid \underline{X}(t_0) = \underline{x}\} = 0(1)$ as $t_1 \to \infty$, then

$$\gamma^* = \liminf_{t_1\to\infty} \frac{E^{\mathcal{P}^*} C(\underline{x},t_0,t_1)}{t_1-t_0} \geq \gamma = \lim_{t_1\to\infty} \frac{E^{\mathcal{P}} C(\underline{x},t_0,t_1)}{(t_1-t_0)}$$

and $\mathcal{P}$ is optimal among the class of stationary policies which satisfy the above restriction.

We apply the optimality condition (5.7) to determine bounds on γ. Suppose that for a given function v^*,

(5.8) $\inf_{\mathcal{P}^*} E^{\mathcal{P}^*}(dv^* + dR) = \Upsilon(\underline{x})dt$

Then r replaced by $r^* = r + \inf_{\underline{x}}\{\gamma(\underline{x})\} - \gamma(\underline{x}) \leq r$,

defines a problem with an optimal expected average cost of $\gamma^* = \inf_{\underline{x}} \gamma(x) \leq \gamma$. From this argument and a similar one involving $\sup_{\underline{x}} \gamma(x)$, we have the bounds

(5.9) $\inf_{\underline{x}} \gamma(\underline{x}) \leq \gamma \leq \sup_{\underline{x}} \gamma(\underline{x})$

As in the deterministic case, the optimality condition (5.7) converts to $|v_{x_2}| \leq c$ on A_0, $v_{x_2} \geq c$ on A_- , and $v_{x_2} \leq -c$ on A_+.

For a given stationary policy $\mathcal{P}$, determined by a specified A_0, A_-, A_+, the potential function is a solution of the partial differential equations (5.3'). The problem of finding the optimal policy is related to the free boundary problem (FBP) of solving the differential equation and finding the regions A_0, A_-, A_+ for which the optimality conditions are satisfied.

In this paper we bypass this analytic problem and consider instead a numerical approximation to the solution of the stochastic control problem by solving a discrete bounded space, discrete time approximation to the problem.

6. Discrete Approximation to the Stochastic Control Problem.

Here we propose to approximate our control problem by a discrete time, finite space Markov Decision problem and to solve that by backward induction.

The laws of motion of the satellite can be approximated by

$$x_1(t+1) = x_1(t) + x_2(t)$$

(6.1)

$$x_2(t+1) = x_2(t) + u(\underline{x}(t),t) + \sigma y(t)$$

where $y(t) = \pm 1$ with probability $1/2$ and u is the control. If u is confined to $\pm u_0$ or 0 where u_0 is an integer and. σ is an integer, a point $\underline{x}(t) = (x_1(t), x_2(t))$ whose coordinates are integers will move to another such point. To bound the state space we must confine $\underline{x}(t)$ to a finite set $\mathcal{E}$ of such points but then $\underline{x}(t+1)$ may no longer be in $\mathcal{E}$. To handle that case we may replace each ordinary successor $\underline{x}(t+1)$ of a point $\underline{x}(t) \in \mathcal{E}$ by a suitably modified successor in $\mathcal{E}$ if $\underline{x}(t+1)$ is not itself a point in $\mathcal{E}$. As a result we will have a modified set of laws of motion where $\underline{X}(t) = \underline{x} \in \mathcal{E}$ and u determine a probability distribution for $\underline{X}(t+1) \in \mathcal{E}$.

We now define the cost associated with the time interval $(t,t+1]$ to be $\tilde{r}(t+1) = r(\underline{x}(t), u(\underline{x}(t),t))$ if the ordinary successor of $\underline{x}(t)$ is in $\mathcal{E}$. Later we shall modify r somewhat. If $\mathcal{E}$ is a set containing all points in a large circle about the origin, one would expect the problem of minimizing the expected long run average cost for this problem to resemble that of our continuous time continuous unbounded state problem. But this discrete time finite space problem can be solved, and backward induction provides a method of approximating its solution.

This program faces a few difficulties. First, since t and the coordinates of x change by integer values, our approximation may be rather coarse and require refinement. Also the values of u_0 and σ may not be integers. Second we have not yet specified $\mathcal{E}$, the modified successor rule, nor $\tilde{r}(t+1)$ if the ordinary successor is not in $\mathcal{E}$. The procedure for specifying the successor and $\tilde{r}(t+1)$ will determine how large $\mathcal{E}$ must be to reduce the edge effects of bounding the state space. A poor choice will lead to large edge effects and require a correspondingly large $\mathcal{E}$ to reduce these effects. But a large $\mathcal{E}$ requires correspondingly more computing effort. Third, backward induction is simple to implement but may require considerable computing. It is possible to reduce the amount of computing needed by having a good approximation to the solution and by using acceleration techniques.

Having described the approach in principle we provide some details. Given a discrete time finite state stationary problem and a terminal cost $v_0(\underline{x})$, let

$$C(\underline{x},t_0,t_1) = \sum_{t=t_0+1}^{t_1} \tilde{r}(t) \tag{6.2}$$

represent the cost over $(t_0 t_1]$ for a procedure given a starting point $\underline{X}(t_0) = \underline{x}$ and let

$$D_{v_0}^{\mathcal{P}}(\underline{x},t_0,t_1) = E^{\mathcal{P}}\{C(\underline{x},t_0,t_1) + v_0[X(t_1)] \mid X(t) = \underline{x}\}$$

be the expected cost for the problem with terminal cost v_0. Backward induction permits one to compute both the optimal average for the problem with terminal cost v_0 and the optimal policy, using the equation

$$D_{v_0}^{\mathcal{P}}(\underline{x},t_0,t_1) = \inf_{u=\pm u_0,0} E[D_{v_0}^{\mathcal{P}}(\underline{X}(t_0+1),t_0+1,t_1)+\tilde{r}(t_0+1) \mid \underline{X}(t_0) = \underline{x}, u] \quad t_0 \le t_1$$

(6.3)

$$D_{v_0}^{\mathcal{P}}(\underline{x},t_1,t_1) = v_0(\underline{x})$$

where we recall that the distribution of $\underline{X}(t_0+1)$ depends on u. The conditional expectation in the above expression is the average of 2 terms involving $y = \pm 1$. Under suitable conditions, concerning non-periodicity and ergodicity, it is possible to show that as $t_1-t_0 \to \infty$,

(6.4) $D_{v_0}^{\mathcal{P}}(\underline{x},t_0,t_1) = \gamma(t_1-t_0) + v(\underline{x}) + o(1)$

where $v(\underline{x})$ is the potential function and γ is the expected long run average cost for the optimal long run stationary policy. Moreover the policy $\mathcal{P}$ also converges in the sense that for $t_1 - t_0$ sufficiently large the backward induction policy at t_0 coincides with the optimal long run stationary policy. Incidentally the closer v_0 is to v, the quicker this method converges.

If we let $D_{v_0}^{\mathcal{P}}(\underline{0},-n,0) - D_{v_0}^{\mathcal{P}}(\underline{0},-(n-1),0) = \gamma_n$ and $D_{v_0}^{\mathcal{P}}(\underline{x},-n,0) - D_{v_0}^{\mathcal{P}}(\underline{0},-n,0) = v_n(\underline{x})$, then $\gamma_n \to \gamma$ and $v_n(\underline{x}) \to v(\underline{x}) - v(\underline{0})$. Substituting $D_{v_0}^{\mathcal{P}}(\underline{x},-(n-1),0)$ for v^* in a discrete version of (5.8) we have $v_n(\underline{x}) - v_{n-1}(\underline{x}) + \gamma_n$ in place of $\gamma(\underline{x})$. Hence bounds on γ are provided by

$$(6.5)\qquad \gamma_n + \inf_x\,[v_n(\underline{x}) - v_{n-1}(\underline{x})] \le \gamma \le \gamma_n + \sup_x\,[v_n(\underline{x}) - v_{n-1}(\underline{x})]\ .$$

Finally $v_n(\underline{x})$ satisfies the slightly simpler looking version of (6.3) which is given by

$$(6.6)\qquad \gamma_n + v_n(\underline{x}) = \inf_{u=\pm u_0,0} E[v_{n-1}[\underline{X}(t_0+1)] + r(t_0+1)\,|\,\underline{X}(t_0) = \underline{x}] \quad n \ge 1$$

where γ_n is defined to be the right hand side of (6.6) when $\underline{x} = \underline{0}$. (Thus $v_n(\underline{0}) = 0$).

In summary once we have a finite stationary Markov Decision problem, the equations (6.5)-(6.6) describe how to compute the optimal γ , v and $\mathcal{P}$, and bounds on these. In the next sections we describe some alternate approaches to handling the difficulties listed above.

7. Refinement of Grid

In Section 4 we discussed the transformation which leaves the problem invariant except for changes in the parameters u_0 , σ and c . If u_0 and σ are replaced by u_0^* and σ^*, c is replaced by c^* and the change of 1 in x_1^* , x_2^* and t^* correspond to changes of a_1^{-1} , a_2^{-1} and a_3^{-1} in x_1, x_2 and t . Thus by taking σ^* and u_0^* to be integers so that σ^*/u_0^* is large, we have fine grids in all 3 scales. For example if $\sigma = u_0 = 1$ and $\sigma^* = 2$ and $u_0^* = 1$, the changes of 1 in x_1^* , x_2^* and t^* correspond to changes of 1/16 , 1/4 , and 1/4 in x_1,x_2 and t .

8. The Finite Set $\mathcal{E}$ of States

If we regard $\mathcal{E}$ as representing a region in the (x_1,x_2) space, some point near the boundary will have a tendency to be followed by a successor not in $\mathcal{E}$. In the infinite set discrete space version of our problem such a point will tend to trace out a path which may be regarded as a temporary

excursion from $\mathcal{E}$. Our strategy is to select $\mathcal{E}$ so that a good deterministic policy would lead to relatively short excursions. It seems intuitively clear that this strategy would be helpful in minimizing the edge effects and making it easy to cope with them. Informal analysis suggested an ellipse centered at the origin which would have its major axis go roughly through the points $(s^2/2u_0,-s)$ and $(-s^2/2u_0,s)$ where s is a size parameter and the designated points are on the parabola leading to 0 in a reasonable suboptimal policy for the deterministic problem. Indeed, the ellipse selected is

$$\frac{x_1^2}{\tilde{a}_1^2} + \frac{x_2^2}{\tilde{a}_2^2} + 2\rho \frac{x_1 x_2}{\tilde{a}_1 \tilde{a}_2} = 1 ,$$

where $\tilde{a}_1 = 1.45s^2/(4u_0)$, $\tilde{a}_2 = 1.7s/2$, and

$$\rho = \frac{(1.45)(1.7)}{8}\left[\frac{4}{(1.45)^2} + \frac{4}{(1.7)^2} - 1\right] = .705 .$$

This ellipse goes through the points indicated above. Note that the area inside the ellipse is proportional to s^3. Thus as s increases and the grid becomes refined, the number of points in $\mathcal{E}$ grows rapidly.

Related to the lattice points inside the ellipse, a

set of points $\mathcal{B}$ called the boundary is identified. These are the possible successors $(x_1+x_2, x_2+u_0 i+\sigma j)$, (where $i = \pm 1$ or 0 and $j = \pm 1$), which are not in $\mathcal{E}$.

9. Edge Effects Adjustment

An initial conjecture that subsequently proved wrong was that differences in the (optimal) potential function for two successive states far from the origin would resemble the differences in the total cost function for the optimal policy in the deterministic problem and that these differences in turn would resemble those for the total cost function for one of the suboptimal policies in the deterministic problem. This conjecture led to the following choice of v_0 and edge effect adjustment. The function v_0 was the total cost function for the relatively easily computed policy which was described in Section 2.

The edge effect adjustment may be described as follows. We shall first present a policy which resembles a good policy for the deterministic problem when $\underline{x}$ is far from the origin. In the unbounded discrete space version of the deterministic problem, this policy would lead from a point $\underline{y} \in \mathcal{B}$ (the boundary) to an excursion which ultimately returns to some reentry point $\underline{x}^* \in \mathcal{E}$ after $m(\underline{y})$ steps. We shall act as though

$$(9.1) \qquad v_n(\underline{y}) = v_n(\underline{x}^*) + [v_0(\underline{y}) - v_0(\underline{x}^*)] - m(\underline{y})\gamma_n \quad .$$

This permits one to apply equation (6.6) even for points $\underline{x} \in \mathcal{E}$ with possible successors in $\mathcal{B}$. In effect we are simply replacing the possible successors $\underline{y}$ of $\underline{x}$ by $\underline{x}^* \in \mathcal{E}$ and changing the cost of using u when $\underline{X}(t) = x$.

Here $v_0(\underline{y}) - v_0(\underline{x}^*)$ is an estimate of the cumulated cost of moving from $\underline{y}$ to x^* and $m(\underline{y})\gamma_n$ is a correction for the average cost of taking $m(\underline{y})$ steps. The latter correction is based on a natural interpretation of (6.6) after transposing γ_n to the right hand side. The above mentioned policy used to determine the excursion and the reentry point $\underline{x}^*$ is to use $u = -u_0$ as long as $\underline{x}$ is such that $x_2 > 0$ and $x_1 > -x_2{}^2/(2u_0)$ or $x_2 < 0$ and $x_1 > x_2{}^2/(2u_0)$, unless this gives a successor (x_1+x_2, x_2-u_0) which overshoots the parabola; i.e., $x_2-u_0 < 0$ and $(x_1+x_2) < -(x_2-u_0)^2/2u_0$. In that case use $u = 0$ instead of $u = -u_0$. This covers half of the $\underline{x}$ space and the other half is treated symmetrically.

At this point we comment that instead of using $r(\underline{x},u) = x_1{}^2 + c|u|$ we use $x_1{}^2 + x_1x_2 + x_2{}^2/3 + c|u|$ on the ground that if $\underline{x}(0) = \underline{x}$ and $dx_1 = x_2 dt$, $\int x_1'^2 dt' = x_1{}^2 + x_1x_2 + x_2{}^2/3$. Thus we would expect the revised $r(\underline{x},u)$ to better reflect the cost of the continuous time problem than the original $r(\underline{x},u)$.

Differences in v_0 between successive points tend to underestimate the corresponding differences in the potential function for the less favorable stochastic problem. Hence this edge effect adjustment tends to lead to a solution with a lower value of γ than that of the infinite discrete stochastic problem. Thus as the size s increases the corresponding values of γ increase. For example with $u_0 = \sigma = c = 1$, we have γ as a function of s in Table 1.

A slight improvement is obtained if $v_0(\underline{y})-v_0(\underline{x}^*)$ in (9.1) is replaced by the sum of the $r(\underline{x},u)$ incurred over the excursion from y to x^*. This replacement yields a better indication of what the edge effect in the discrete approximation should be for larger s. The improvement is reflected in that for a given vlaue of s, the resulting long run expected average cost γ^* for the revised version is slightly closer to the limiting value. See Table 1.

One would expect the values of γ to decrease as we use more refined grids. Indeed Table 2 presents some values of γ^* for 3 grid size parameters which are progressively more refined. The latter require many more points for a given s and only relatively small values of s were treated in preliminary calculations. In later calculations, the expectation of decreasing γ is not realized. A possible explanation is conjectured in conclusion (c) in Section 11.

Note that we must deal with a triple limit. The number of iteration n must $\to \infty$, the size of the ellipse s must $\to \infty$ and the grid sizes must $\to 0$.

Several alternatives were employed to improve the edge effect performance so that good approximations could be achieved without an undue computing burden. These are described in the next Section.

10. Alternative Edge Effect Adjustments

Two major alternative approaches were used. One was to start with coarse grids with large size ellipses. From the iterations in this case, v is estimated by interpolation inside the ellipse. These estimated values of v were used to help estimate edge effect adjustments in later computations with finer grids and smaller ellipses. The second approach was to simulate random excursions from $\mathcal{E}$ to estimate v on the boundary. We present more detail below.

(a) Interpolation

Suppose that the approach of Section 9 has been applied for a certain grid and a large ellipse $\mathcal{E}_1$ of size s_1. After a number of iterations, good estimates γ and v and the optimal policy are obtained for the corresponding discrete approximation to our problem. Select a finer grid (by choosing new values of u_0^* and σ^*) and a correspondingly smaller

ellipse $\mathcal{E}_2$ of size s_2 to keep the number of points in $\mathcal{E}_2$ within bounds. By interpolation from v compute v^* the estimated values of v on the new grid in the new ellipse $\mathcal{E}_2$ and on its boundary $\mathcal{B}_2$. For each point $\underline{x} \varepsilon \mathcal{E}_2$ which has a possible successor $\underline{y} \varepsilon \mathcal{B}_2$ define $d(\underline{x},\underline{y}) = v^*(\underline{y}) - v^*(\underline{x})$. Hereafter, in doing the backward induction the term $v_n(\underline{y})$ in the computation of $E\{v_n(\underline{X}(t+1)) | \underline{X}(t)=\underline{x}\}$ is replaced by $v_{n-1}(\underline{x}) + d(\underline{x},\underline{y})$. With this treatment of the edge effect, apply backward induction until good estimates of γ and v are obtained for the new grid size. This process of refinement of grid and reduction of size can be repeated using the interpolation technique.

Simply reducing the size of the ellipse without changing the grid size shows how stable the method is. We find for example that with $u_0 = c = \sigma = 1$ and $s = 12.0$ $\gamma = 9.2626$. Successive reductions in sizes from $s = 12.0$ to $s = 6.4$ and $s = 2.4$, without refinement in grid, lead to estimates which require no changes of γ and v through further iteration. Such excellent results cannot be expected when the grid is refined for then the refined discrete problem should have a somewhat different answer depending on how coarse the original grid was. Another potential difficulty is that the interpolation process is not very accurate for v on a coarse grid. Indeed the bevahior of v as x_2 changes is difficult to approximate well by linear or quadratic interpolation. Although the estimated values within $\mathcal{E}$ are not crucial, the values of $d(\underline{x},\underline{y})$ are very important in determining

the limiting values of γ and v, especially when the size s of the ellipse is small.

In this reduction and interpolation approach the Markov Decision Problem with finite state space $\mathcal{E}_2$ has been replaced by a new problem where $\mathcal{E}_2$ is augmented by the points $\underline{y} \varepsilon \mathcal{B}_2$. However if $\underline{y} \varepsilon \mathcal{B}_2$ is a successor of two distinct $\underline{x} \varepsilon \mathcal{E}_2$, the augmented state space must treat $\underline{y}$ as two points, else there will be a discrepancy due to the fact that $v_{n-1}(\underline{x}) + d(\underline{x},\underline{y})$ may not coincide for both $\underline{x}$. In this related problem the equation $v_n(\underline{y}) = v_{n-1}(\underline{x}) + d(\underline{x},\underline{y})$ implies a motion from $\underline{y}$ to its predecessor $\underline{x}$ with a related cost $d(\underline{x},\underline{y}) + \gamma_n$ which is almost stationary.

(b) Random Excursion.

The random excursion edge effect adjustment differs from the effect described in Section 9. There we modeled a problem which is essentially one where points outside the ellipse travel without the influence of random forces, and are subjected to the suboptimal deterministic policy. This problem underestimates the cost of excursions from $\mathcal{E}$. A more realistic estimate can be obtained by simulating the motion with the random forces until the point which has left $\mathcal{E}$ returns to $\mathcal{E}$. To do so, a point which leaves $\mathcal{E}$ from $\underline{x}$ moves to $\underline{y} \varepsilon \mathcal{B}$ and from $\underline{y}$ to $\underline{y}' = (y_1+y_2, y_2 + u + \sigma w)$ where u is selected to be u_0, 0 or $-u_0$ according to the optimal deterministic policy and w

is selected to be $+1$ or -1 with probability $1/2$ by use of a random number generator. From $\underline{y}'$ the point moves on until it ultimately returns to $\mathcal{E}$ at a point $\underline{x}^{(1)}$ after incurring a cumulated cost $R^{(1)}(y)$ in $n^{(1)}(y)$ steps from $\underline{y}$. This process is repeated m times. Hereafter $v_n(\underline{y})$ is replaced by

$$\frac{1}{m} \sum_{i=1}^{m} \{ v_n(\underline{x}^{(i)}) + R^{(i)}(y) - n^{(i)}(\underline{y}) \gamma_n \}$$

In effect we have a Markov decision problem where the state $\underline{y}$ has been replaced by m equally likely states in $\mathcal{E}$ with appropriate costs attached to the motion required to reenter $\mathcal{E}$.

This edge effect treatment tends to overestimate the cost of excursions in the discrete problem since the policy followed outside $\mathcal{E}$ is suboptimal. This tendency is relatively slight and good estimates of γ can be obtained for smaller ellipses than those used in Section 9. However these estimates are affected by the random process used in the simulation and fluctuate from simulation to simulation. Table 4 presents the results of several such simulations. This indicates that the simulation techniques is quite effective when s is as large as 6. For smaller s, there is a great deal of variability and the fact that the excursions are guided by a suboptimal policy introduces a positive bias.

An elaboration of this approach is to apply the method of importance sampling. Here one decides for each $y \notin \mathcal{E}$ which value of w, plus or minus one is most favorable in the sense that it more resembles the optimal deterministic policy. The faborable value of w is selected with probability less than 1/2. This biased sampling procedure gives distorted estimates which are easily compensated for by the methods of importance sampling. To date limited experimentation with importance sampling has not shown great improvement over the simpler simulation although such techniques are often good for reducing variance.

The combined effect of reducing size followed by random excursion simulation was tried with a reduction from $s = 10$ with $u_0^* = \sigma^* = 1$ to $s = 3$ and 4 with $u_0^* = 2$, $\sigma^* = 3$. The results were variable and were not as good as using the simpler reduction plan. The conclusion seems to be that random excursion simulations should be avoided unless s is large (greater than 6 for $u_0 = \sigma = c = 1$).

1. Miscellaneous Remarks

(a) Acceleration Techniques

The study of successive values of γ_n indicates that after an early period of major adjustments γ_n tends to fluctuate periodically about the limiting value. In particular for $c = u_0 = \sigma = 1$, successive values alternate below and above the limit. In that case the occasional replacement of $v_n(\underline{x})$ by $[v_n(\underline{x}) + v_{n-1}(\underline{x})]/2$ and γ_n by $(\gamma_n + \gamma_{n-1})/2$ accelerates the convergence. On other occasions the use of $(\gamma_{n-2} + 2\gamma_{n-1} + \gamma_n)/4$ and a similar operation on v_n proves helpful. In cases where γ_n seems to be increasing steadily by small almost equal increments, the occasional use of $\gamma_n + a(\gamma_n - \gamma_{n-1})$ for $a>0$ speeds up convergence. Without these acceleration techniques the case of $u_0 = c = \sigma = 1$, $s = 9.0$ required $n = 150$ to converge to the point where the $\sup_{\underline{x}} |v_n(x) - v_{n-1}(x)| \leq .00024$. With 3 applications of these simple acceleration techniques, only 70 iterations were needed to obtain this result. These averaging methods are related to what Kushner and Kleinman call the accelerated Jaboci Method [8].

(b) Evaluation of Suboptimal Policies.

A major function of finding and evaluating optimal policies is to decide whether a convenient or simple suboptimal policy is relatively efficient. To do so one must also be able to evaluate a specified suboptimal policy. The general policies described in this paper are easily adapted to the problem of

evaluating a specified stationary policy. The fundamental equation (6.6) is changed so that the infimum is omitted and the distribution of $\underline{X}(t_0+1)$ is governed by the specified policy.

(c) The Method of Kushner and Kleinman

Kushner and Kleinman [7] and Kushner [6] present an approach very similar to that of this paper in a related problem. However their treatment of the edge effect was simpler and somewhat less inclined to give good approximations for a given size region. First the region is rectangular rather than elliptical. Second, points on the boundary of the rectangle are constrained to move along the boundary when they do not reenter naturally. In effect the boundary acts much like a reflecting barrier in the treatment of Kushner and Kleinman. This is a less realistic model than that obtained from our treatment of simulated deterministic or random excursions.

(d) Rigorous Treatment.

A more rigorous treatment of the original continuous time continuous unbounded state space problem involves several difficulties. One is measure theoretic in nature but that seems to be subject to treatment. For example see Yamada [9]. Another problem is that caused by the unboundedness of the state space and the potential function. This seems to be a more fundamental difficulty even in discrete space problems. Recent approaches to these problems have been made by Bather [2] and by Hordijk, Schweitzer and Tyms [4].

11. Computation

After a series of experiments with various approaches, one long computer run was executed in an interactive mode. This run applied a sequence or reductions in size s combined with interpolations to more refined grid sizes. Table 5 outlines some of the details of this run carried out for $u_0 = \sigma = c = 1.0$.

In more detail, the initial approximation to v is v_0 derived from the suboptimal deterministic policy of Section 2. The results of the successive stages were potential functions labeled $v(s,u_0^*,\sigma^*)$ each of which was interpolated to serve as an initial approximation for the next stage. One should recall that the essential aspect of these approximations are their influence on the edge effect. Changing the approximation inside the ellipse only affects the speed of convergence, and that, only to a relatively minor degree.

Table 6 contains these estimates of the poential function and also $v(10,1,1)$ or $v(16,1,1)$. These are based on one stage starting from v_0. The latter, $v(16,1,1)$ was inserted when it was available and $v(10,1,1)$ was not. The table has a considerable number of gaps due partly to incomplete print out detail but mainly to essential unavailability. However, the data, as presented, permit many comparisions to be made and from these a reasonable picture of the nature of the potential functions and the edge effects may be recovered.

Figure 2 presents an approximation to A_0. derived from this run. Figure 3 presents, on a larger scale, the coarse approximation to A_0 derived from the first stage of the run. Several conclusions may be drawn.

(a) More interesting figures for A_0 would have resulted if a larger value of c had been selected. Then the A_0 region would have been larger and the essential coarseness of the grid would have been relatively less important.

(b) A few iterations provides a good estimate of A_0. A coarse grid and an ellipse with relatively few points provides a good rough approximation to γ and v with little computing effort. Refinement by this backward induction technique quickly becomes very expensive.

(c) At first, refinement of grid size provides the controller more opportunity to control properly and γ is reduced. Subsequently another effect tends to increase γ in later stages. While inaccurate interpolation may contribute a deleterious edge effect which raises γ, we conjecture that another effect is more important. The Brownian Motion was modeled by a random variable which takes on values ± 1. This model for a short time grid interval more nearly resembles Brownian Motion over a unit time interval and hence has a higher fourth moment. Thus the ± 1 model for coarse grids has a tendency to reduce the resulting γ in comparison with the Brownian Motion model. An additional byproduct of this effect seems to be that somewhat larger values of s are required as the grid becomes refined.

Acknowledgements

The authors wish to thank Dr. L. E. Shepp for stimulating and helpful discussions on the deterministic problem. We wish to thank David Edelman for the considerable effort and assistance he gave in the writing of the fortran program for the computation.

References

[1] Bather, J. A., "Control charts and the minimization of costs (with discussion)," J. Roy. Stat. Soc., Series B, 25 (1963) 49-80.

[2] Bather, J. A., "Optimal stationary policies for denumerable Markov chains in continuous time", Adv. Appl. Prob., 8 (1976) 144-158.

[3] Dynkin, E. B. Markov Processes, Volume 1, Academic Press, New York, N.Y. (1965).

[4] Hordijk,A. Schweitzer, P. J., and Tijms, H., "The asymptotic behaviour of the minimal total expected cost for the denumerable state Markov decision model", J. Appl. Prob., (1975) 298-305.

[5] Howard, R. A. Dynamic Programming and Markov Processes, M.I.T. Press, Cambridge, Mass. (1960).

[6] Kushner, H. Introduction to Stochastic Control, Holt, Rinehart and Winston, Inc., New York, N.Y. (1971).

[7] Kushner, H. J. and Kleinman, A. J. "Numerical methods for the solution of the degenerate nonlinear elliptic equations arising in optimal control theory", IEEE Trans. Automat. Contr. AC-13 (1968) 344-353.

[8] Kushner, H. J. and Kleinman, A. J., "Accelerated procedure for the solution of discrete Markov control problems", IEEE Trans. Automat. Contr. AC-16 (1971) 147-152.

[9] Yamada, K., "Discrete-time approximation of Markovian control systems", Tech. Rept. Japan UNIVAC Research Inst., (1976) 1-39.

Table 1

Long run expected average costs γ and γ^* for optimal policy in discrete approximation problems as a function of the size parameter s

$c = u_0 = \sigma = 1 \qquad u_0^* = \sigma^* = 1$

s	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0
γ	4.286	6.457	8.370	8.957	9.156	9.230	9.255	9.261	9.262	9.2625	9.2626
γ^*	5.522	7.327	8.666	9.043	9.190	9.241	9.258	9.261	9.262	9.2626	9.2926

Table 2

Long run expected average cost γ^* for optimal policy in discrete approximation problem as a function of size s and grid parameters (u_0^*, σ^*)

$c = u_0 = \sigma = 1$

$s \backslash (u_0^*, \sigma^*)$	(1,1)	(2,3)	(1,2)
3	5.52	3.49	3.05
4.0	7.33	4.92	4.41

Table 3

Limiting values γ^* without reduction and γ after reduction from $s = 10$; $c = u_0 = \sigma = 1$

$s \setminus (u_0^*, \sigma^*)$		(1,1)	(2,3)	(1,2)
3.0	γ^*	5.52	3.49	3.05
	γ	9.26	7.84	8.12
4.0	γ^*	7.33	4.92	4.41
	γ	9.26	7.78	7.96

Table 4

Mean $\bar{\gamma}$ and standard deviation s_γ based on samples of 9 trials with 40 excursions. $u_0 = \sigma = c = 1$, $u_0^* = \sigma^* = 1$

s	4.0	6.0	8.0
$\bar{\gamma}$	10.34	9.26	9.26
s_γ	.64	.097	.004

<u>Table 5</u>

Details of Computer Run $u_0 = \sigma = c = 1.0$

i	s	u_0^*	σ^*	Δx_1	Δx_2	Δt	$n(\mathcal{E})$	$n(\mathcal{B})$	γ
1	20.0	1	1	1.0000	1.0000	1.0000	5,455	928	9.263
2	9.0	1	2	0.0625	0.2500	0.2500	31,863	4,243	7.254
3	4.0	1	3	0.0123	0.1111	0.1111	31,863	5,463	7.471
4	2.2	1	4	0.0039	0.0625	0.0625	29,788	6,403	7.597

i = stage; s = size; $\Delta x_1, \Delta x_2, \Delta t$ are grid sizes in x_1, x_2, t scales; $n(\mathcal{E})$ and $n(\mathcal{B})$ are the number of lattice points in the i-th ellipse and i-th boundary. γ = the long run average cost for the i-th stage approximation to the continuous optimization problem.

Table 6

Approximations to Potential Function

$u_0 = \sigma = c = 1$

v_0	v(9,1,2)
v(20,1,1)	v(4,1,3)
v(10,1,1)	v(22,1,4)

x_2 \ x_1	0.0		0.5		1.0		1.5		2.0	
8.0	8829 20964 20624'									
6.0	2105 6393 5569	5789	6649*	6027	6911	6273	7181*	6530	7460	6798
4.0	285 1319 1286	1116	330 1416*	1201	379 1518	1292	433 1627*	1387	491 1743	1487
2.0	13 124 124	98 98	20 145*	117 117	28 167	138 138	40 195*	163 162	53 228	190 180
1.5	5 53*	40 41 41	8 69*	51 52	14 86*	65 65	20 107*	81 81	29 131*	100 99
1.0	1.7 9.3 9.3	12.4 12.9 13.2	3.5 19.2	18.1 18.6 18.8	7 33	26 26	11 49	36 36	17 67	49 48
0.5	0.6 2.3*	1.8 2.3 2.3	1.5 8.5*	4.4 4.7 4.3	3.3 18.9*	8.5 8.8 8.2	6 32*	15 15	10 49*	22 22
0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.6 2.2*	0.9 0.6 0.6	1.8 8.9	2.5 2.6 2.7	3.7 19.3*	5.7 6.1 6.1	6 34	11 11
-0.5	0.6 -8.5*	1.8 2.3 2.3	0.6 -9.4*	0.9 1.4 1.5	1.2 -5.6*	2.0 2.2 2.3	2.6 2.2*	4.4 4.4 4.4	4.6 14.2*	7.9 7.9 7.9
-1.0	1.7 9.3 9.3	12.4 12.9 13.2	1.1 3.3*	8.8 9.3 9.5	1.3 1.3	6.8 7.5 7.8	2.1 4.1*	6.8 7.5 7.8	3.7 11.3	8.5 9.3 9.5
-1.5	5 38*	40 41 41	3 26*	31 32 33	2 14*	25 26 26	2 9*	21 22 23	3 9*	19 20 21
-2.0	13 124 124	98 98	8 102*	82 83 83	5 80	69 70 70	3 65*	58 60 60	4 55	51 52 53
-4.0	285 1319 1286	1116	244 1226*	1037	206 1137	962	173 1055*	891	143 979	826
-6.0	2105 6393 5569	5789	6144*	5537	5902	5296	5668*	5070	5441	4852'
-8.0	8829 20964 20624'									

* interpolated value

' v(16,1,1) in place of v(10,1,1)

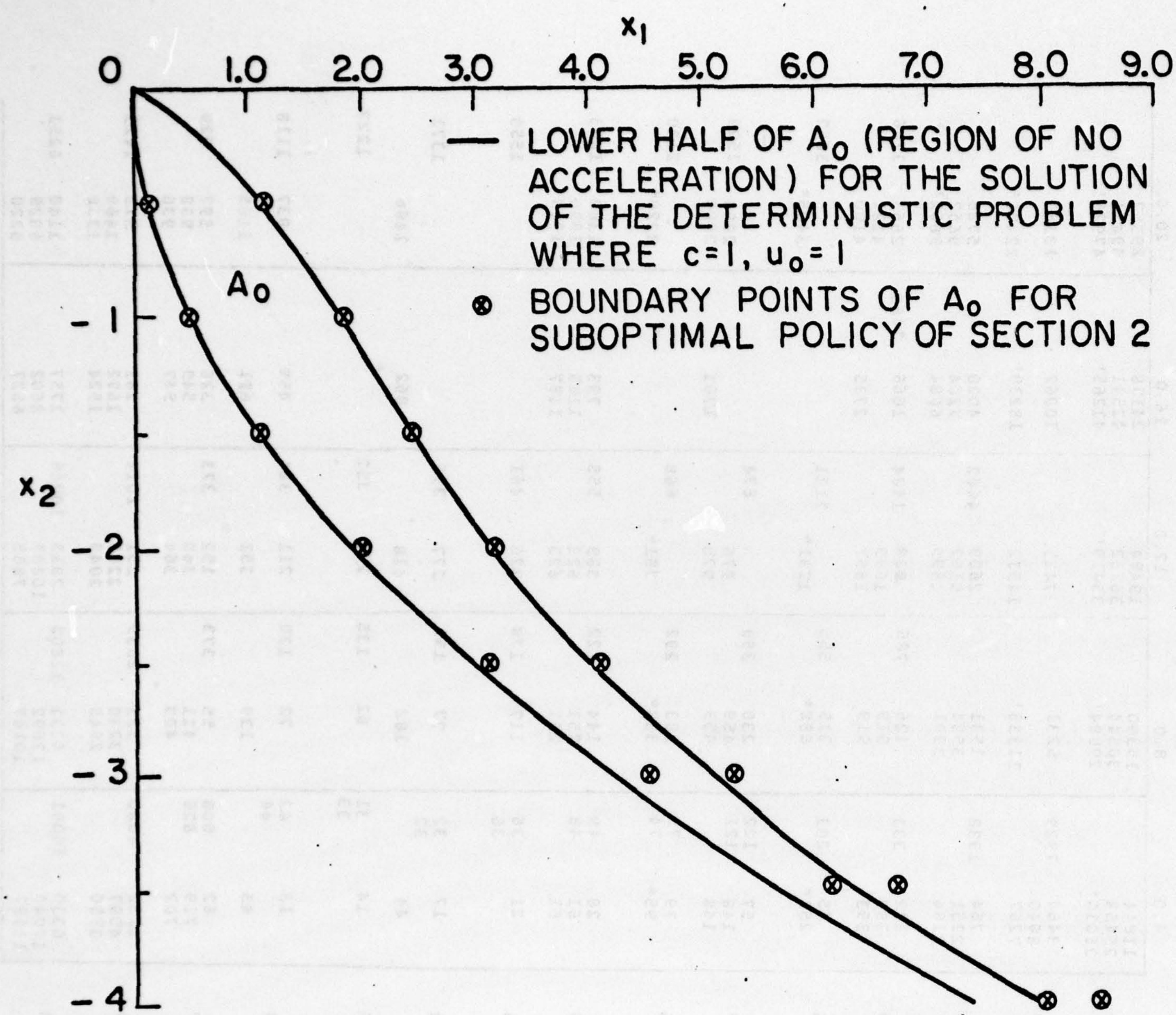
Table 6 (continued)

Approximations to Potential Function

$u_0 = \sigma = c = 1$

v_0	v(9,1,2)
v(20,1,1)	v(4,1,3)
v(10,1,1)	v(22,1,4)

x_2 \ x_1	4.0		8.0		12.0		16.0		20.0	
8.0	11844		15390		19484		24138		29367	
	25458		30544		36232		42541		49474	
	25036'		29884'		35419'		41265'		47926'	
6.0	3464	7929	5231		7425		10062		13160	
	8640									
	7267		11338'		14532'		18230'		22406'	
4.0	764	1938	1531		2609	4662	4020		5785	
	2251		3524		5162		7204		9652	
	2184		3391		4899		6684		9652'	
2.0	132	333	426	785	928	1484	1666	2460	2663	3746
	393		919		1699				4121	
	393		919		1697		2735		4100	
1.5	85	202	315	556		1131				3062
	251*		658*		1291*				3405*	
1.0	57	122	238	398		874				
	148	121	459		976				2804	2529
	148		459		975		1763		2800	
0.5	39	75	183	292		668				2120
	95*	74	338*		781*				2328*	
0.0	28	49	144	222	389	555	793		1380	1803
	61	48	251		623		1188		1950	
	61		251		623		1187		1948	
-0.5	21	36	117	178	326	461				1559
		36								
-1.0	17	32	97	151	277	396				1371
		32								
	44		182		418		862		1486	
-1.5	14	31	82	135	240	352				1227
		33								
-2.0	13	43	72	130	211	324	459		837	1118
		46								
	45		139		351		691		1165	
-4.0	62	608	55	373	152	373	326		597	929
	719	825	411		368		549		938	
	702		403		364		547		936	
-6.0	1132	4080	523	2839	251	2014	281		513	1421
	4597		3230		2269		1692		1468	
	3950		2845		2049		1524		1368	
-8.0	6330	16091	4331	12805	2813	10066	1757		1140	6251
	17046		13692		10883		8602		6829	
	11697		10147		7885		6637		5220	



x_1
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
x_2
-1
-2
-3
-4
A_0
LOWER HALF OF A_0 (REGION OF NO ACCELERATION) FOR THE SOLUTION OF THE DETERMINISTIC PROBLEM WHERE $c=1$, $u_0=1$
BOUNDARY POINTS OF A_0 FOR SUBOPTIMAL POLICY OF SECTION 2

FIGURE 1

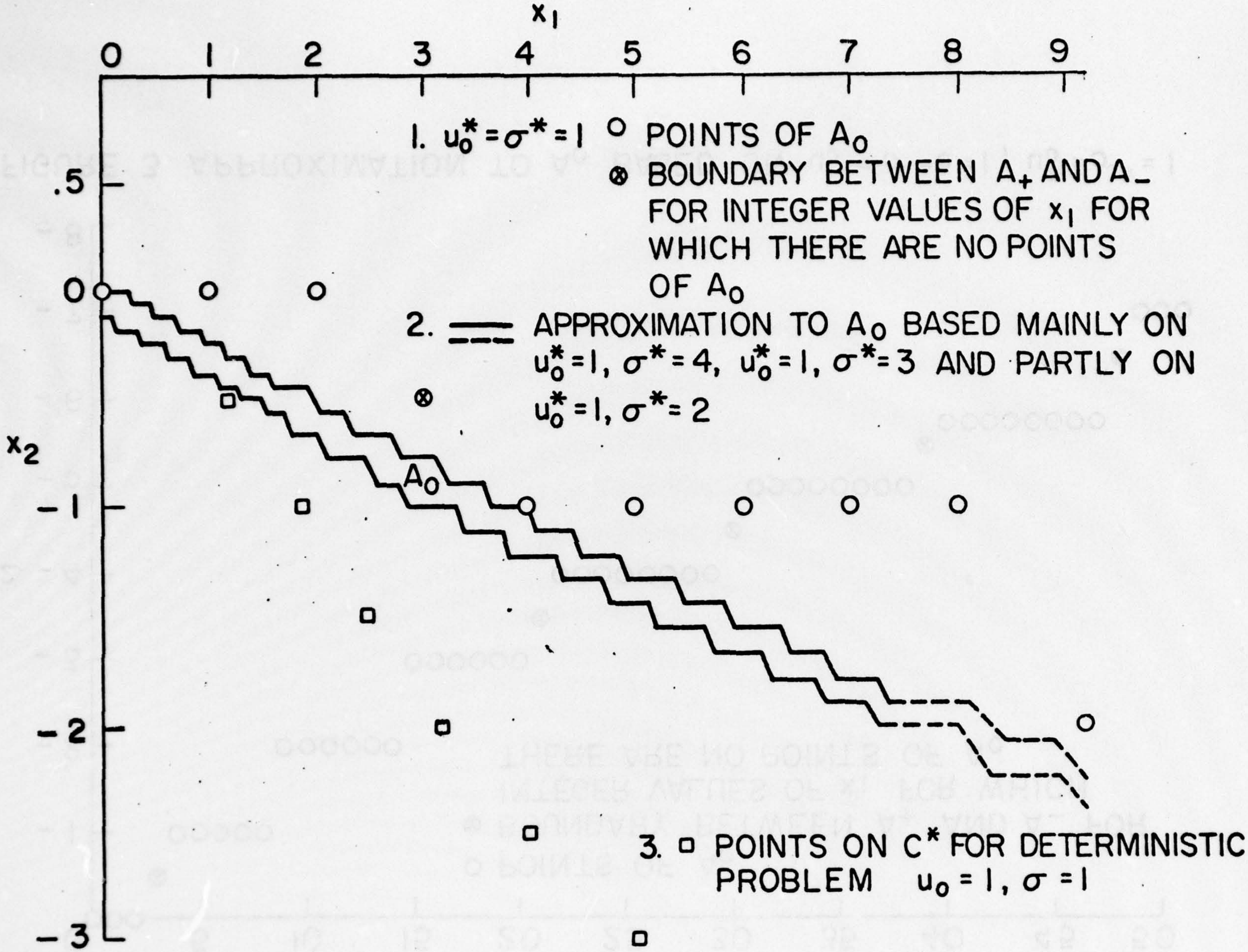


FIGURE 2 APPROXIMATION TO A_0 BASED ON COMPOSITE CALCULATION $u_0 = \sigma = c = 1$

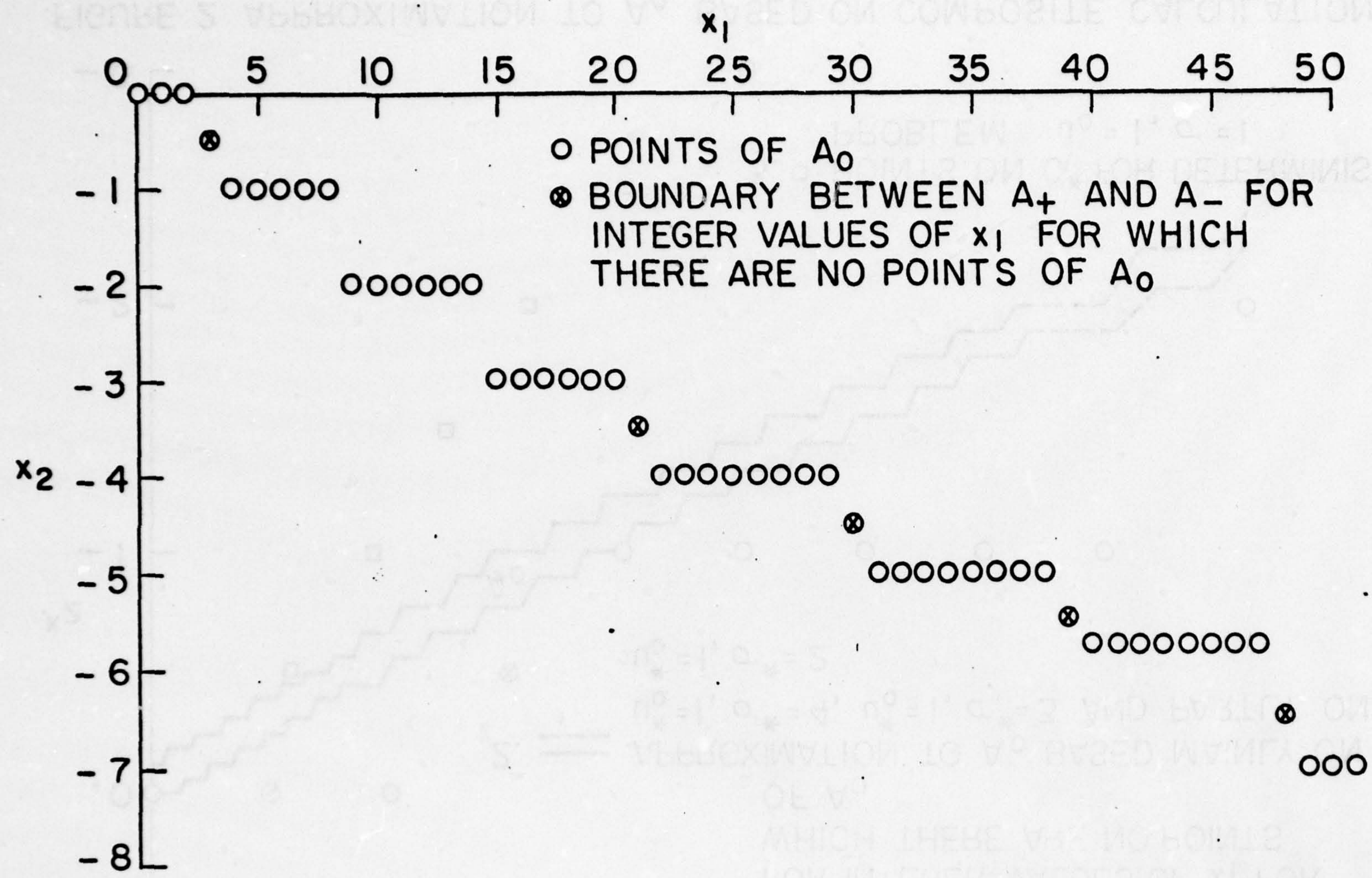


FIGURE 3 APPROXIMATION TO A_0 BASED ON $u_0 = \sigma = c = 1$, $u_0^* = \sigma^* = 1$

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 9	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Satellite Control Problem		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Herman Chernoff A. John Petkau (Univ. of British Columbia, CANADA)		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0555
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematics M.I.T., Cambridge, Mass. 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-331)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Va. 22217		12. REPORT DATE December 22, 1977
		13. NUMBER OF PAGES 52
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Stochastic Control, Dynamic Programming, Markov Decision Problems, Average Cost Criterion, Numerical Approximation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(see reverse side)

over

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

x sub 1 x sub 1 squared

A numerical approach is described for calculating the optimal policy in the stochastic control problem of keeping a satellite close to a fixed point in space when it is subject to random forces. The random forces are modelled by Brownian Motion. A policy is evaluated in terms of its long run expected average cost. The running costs consist of a charge for fuel used plus a charge of x_1^2 per unit of time when the satellite is x_1 units away from the target. The space is one-dimensional. The method used is to apply backward induction to a bounded discrete space, discrete time version of the problem. Incidentally a solution is presented for the deterministic version of the problem where there are no random forces.